



Design of a Multidimensional Model Using Object Oriented Features in UML

Payal Pahwa, Shweta Taneja

Shalini Jain

Design of a Multidimensional Model Using Object Oriented Features in UML

Payal Pahwa

**Principal, Bhagwan Parshuram Institute of Technology,
New Delhi – India**

Shweta Taneja

**Research Scholar, Dept. of Computer Engineering,
Delhi Technological University, New Delhi – India**

Shalini Jain

**Asstt. Professor, Maharaja Surajmal Institute of Technology,
New Delhi – India**

Design of a Multidimensional Model Using Object Oriented Features in UML

ABSTRACT

A data warehouse is a single repository of data which includes data generated from various operational systems. Conceptual modeling is an important concept in the successful design of a data warehouse. The Unified Modeling Language (UML) has become a standard for object modeling during analysis and design steps of software system development. The paper proposes an object oriented approach to model the process of data warehouse design. The hierarchies of each data element can be explicitly defined, thus highlighting the data granularity. We propose a UML multidimensional model using various data sources based on UML schemas. We present a conceptual-level integration framework on diverse UML data sources on which OLAP operations can be performed. Our integration framework takes into account the benefits of UML (its concepts, relationships and extended features) which is more close to the real world and can model even the complex problems easily and accurately. Two steps are involved in our integration framework. The first one is to convert UML schemas into UML class diagrams. The second is to build a multidimensional model from the UML class diagrams.

The paper focuses on the transformations used in the second step. We describe how to represent a multidimensional model using a UML star or snowflake diagram with the help of a case study. To the best of our knowledge, we are the first people to represent a UML snowflake diagram that integrates heterogeneous UML data sources.

Index Terms — *Data warehouse, Multidimensional modelling, Object Oriented Paradigm, Unified Modelling Language.*

INTRODUCTION

A data warehouse is defined [5] as a subject oriented, integrated, time variant, non-volatile, historical collection of data in support of management decision making process. A data warehouse (DW) is used to handle complex, sophisticated, online and multidimensional analysis of data by fetching information from multidimensional databases. For the purpose of data analysis from a DW, the notion of the cube has been widely accepted as the underlying logical structure of DW or multidimensional databases. The DW design is of three types—namely, the conceptual , dealing with the high level representation of the world in order to capture the user ideas using rich set of semantic constructs. The physical design deals with the details of the representation of the

information or data storage techniques in the specific database management system. The logical design acts as an intermediate between the above two mentioned extremes, trying to balance a storage independent paradigm and a natural representation of the information in terms of computer oriented concepts.

The Unified Modelling Language (UML) in has become an industry standard for object modelling during analysis and design steps of software development [3], [13]. We have used UML in our design which is a widely accepted modelling language. So, we need not learn new notations and methodologies for modelling. Another excellent feature of UML is that it is an extensible language, that is, it provides features like stereotypes, tagged values etc. to add new elements to domains like business modelling, web applications etc.

The DWH paradigm involves complex queries on large amounts of data, which are difficult to manage for human analysts. The entity-relationship data model, proposed earlier is appropriate for online transaction processing. A data warehouse, however, requires a concise, subject oriented schema that facilitates online data analysis. The most popular data model for a data warehouse is a multidimensional model. The data is located in n-dimensional space, with the dimensions representing the different ways the data can be viewed. There are various ways in which multidimensional data can be represented like star schema, snowflake schema or a fact constellation schema. A star schema is basically a relational model in the shape of a star. At the centre of the star there is the fact table containing data on the subject of analysis and a set of smaller tables called dimension tables, one for each dimension. The fact table contains keys for each of the dimensions along with numerical attributes called measures or fact attributes. For example, a sales data warehouse has a central fact table called sales. The fact table contains two measures, that is, dollars sold and units sold. Sales are considered along four dimensions namely time, item, branch and location. Multidimensional modelling has two advantages. Firstly the model is more close to the real world and can help the data analysts to model the concepts easily. Secondly, it is simple in structure for the users to understand. The snowflake schema is a variant of the star schema model, where some dimension tables are normalized, thereby further splitting the data into additional tables. A fact constellation schema has more than one fact table and allows dimension tables to be shared between fact tables. We have used snowflake model in our design as the data is in normalized form, which can reduce redundancies and is easy to maintain. The concept of data granularity is also modelled in the paper. The analysis of data can be done from high level to lower levels of detail. Depending on user's query, we can go to particular level of detail and satisfy the query.

The paper is organized as follows. Section 2 presents the related work done in this field. In Section 3, the system architecture has been proposed. Section 4 shows a case study of Diabetic Interactive Electronic Treatment System. A three level system design is presented in section 5. The following section describes the mapping rules made. Finally the conclusion is stated.

Literature Survey

In the past few years, several approaches for representing the main multidimensional (MD) properties at the conceptual level have been proposed in [4], [8] and [11], [12] respectively. Nevertheless, none of these approaches for MD modelling provided a generic model which could be accepted and covered all aspects of a data warehouse. A number of multidimensional models built at the conceptual level have been proposed. The authors in [1] gave a complete comparison of MD data models. In [7], the authors present the process of data warehouse architecture, development and design. They highlight the different aspects to be considered in building a data warehouse. These range from data store characteristics to data modelling and the principles to be considered for effective data warehouse architecture.

In [10], the authors introduce a UML profile for modelling different kinds of DWH usage on a conceptual level. They distinguish four perspectives of usage -access control, temporal intensity, temporal flexibility and importance, as well as active or passive usage .They have basically considered the details of the users such as their skill level, number of instances, functional grouping etc. neglecting the designer's role. The authors in [2] present a UML profile for modelling DWH usage on a conceptual level. It uses features of UML intended for the purpose of creating abstract, general models. The profile distinguishes four perspectives of usage and allows to model details of the users. The UML profile is applied to example illustrating Hajj pilgrim's private tour. In [15], the authors have presented the relation of provider and user for data in various departments of government in detail. On the base of data analysis, the architecture of data centre based on data warehouse technology has been presented. This work was done for Chinese electronic government system.

A Graph based Object Oriented Multidimensional Data model (GOOMD) for the conceptual level design of a data warehouse has been proposed in [9]. None of the approach provides a generic model which can be accepted.

We have designed a conceptual UML snowflake model from various UML data sources. The main advantage of our method is that it is based on a well known standard modelling language. So, the designers can avoid learning a new language for multidimensional systems. To the best of our knowledge, we are the first people to represent a UML snowflake diagram that integrates heterogeneous UML data sources.

System Architecture

The architecture of our proposed system is a 2-tier architecture shown in figure 1.The major tier is the data integration, which translates UML schemas associated with UML

data sources into conceptual-level UML class diagrams and then constructs a multidimensional model based on the UML diagrams.

The second tier is the data query process in which a query issued by user is decomposed into the sub-queries onto the different UML data sources, and then the results retrieved from the individual UML data sources are integrated and returned to the user. The component Query Originator performs the function of decomposing queries into sub-queries and the component Query Integrator performs the function that integrates the results of sub-queries and returns it to the user who issues the operation.

Diabetic Interactive Electronic Treatment System (D.I.E.T) - a Case Study

The Diabetic Interactive Electronic Treatment System aims at doing best service to a patient. It is a computer based interactive program to record and retrieve the patient details such as registration, lab results, symptom and food/exercises to be followed. In simple, it maintains complete history of each patient. The appointment and registration of patient generates a confirmation slip by verifying the doctor's availability. The physical examination is done for the patient. The lab test module suggests the necessary general and lab tests. The advice of food habit, physical exercises to be followed will be chosen in the system and summarized by doctor.

System Design

Based on our experience of the real world problems, we know that complex diagrams are difficult to model and understand, so in our approach we have decomposed the levels of details of the design process into three. We have shown the three levels below by applying them on the Diabetic Interactive Electronic Treatment System (D.I.E.T) case study.

The design of data warehouse can be divided in three levels -

1. Level 1- Conceptual Model.
In this level, we show the conceptual multidimensional model of the entire system in the form of Star/Snowflake Schema.
2. Level 2- Star Schema
This level shows the facts and dimensions of a star model.
3. Level 3-Snowflake Schema
In this level, the dimensions and their hierarchy levels i.e. classification or dimension hierarchy etc. are shown. Here, the dimensions of the star schema and obtained the snowflake schema are normalized. For example, in the following

figure, we have shown the dimension hierarchy of patient dimension already shown above in the star model.

Mapping Rules

The mapping rules specify the generation of conceptual level UML multidimensional model from UML Class Diagram. In a UML conceptual multidimensional model (Figure 6), the fact class and dimension class are made from UML classes. The fact class and dimension class are related using a UML association. The cardinality of two associated class are 0...n (*) and 1 respectively specifying that each dimension object relates to zero, one or more fact objects. The UML classes participating in aggregation property are modelled as a dimension hierarchy in UML multidimensional model. The cardinality of a lower level class to the higher level class of the same hierarchy is 1...n (*) and 1 respectively. The attributes of class diagram can be modelled as dimension attributes or measures, depending on the user's and designer's point of view. According to Moody D.L. et al. (June 2000), the generalizations of the UML notation cannot be mapped directly to hierarchies in the multidimensional model in [6], since the semantics of hierarchies in object-oriented models and multidimensional models differ. We want to preserve the information contained in UML generalizations and transform these hierarchies to enable their correct mapping to multidimensional hierarchies in the logical phase. So, we transform the generalizations into aggregations and classes following the proposal for ER models [14].

UML Class Diagram and Resultant UML Snowflake Model

The UML snowflake diagram is a type of UML conceptual multidimensional model. The process of converting UML class diagrams that come from multiple UML data sources is not simple, but requires the knowledge of the domain expert who knows what information should be included in the resultant UML snowflake diagram. Due to limited space, we only show the resultant UML snowflake diagram from integrating four UML documents. They are Patient document, Hospital document, Doctor Document, and Lab Test document. The patient document describes the ID of the patients, their symptom of disease etc. The hospital document describes the name of hospital; its address (street, city, state, zip and country).The doctor document specifies ID of the doctor, name, address etc. The Lab Test document shows the general and specific tests.

In the UML Class diagram shown below, extended features of UML like generalization, specialization, constraints etc. are used. A constraint may be denoted on paths of association or as a note. Here one of the doctor belongs to the department is head of the department.

Data Granularity models the level of detail present in the data sources. Lower the level of detail, finer be the data granularity. Rolling up and drilling down of data becomes more efficient by defining the data grain at each level of hierarchy. By using object oriented concepts we have defined relationships between various data entities which help us to understand the interdependence between these data entities on every level of hierarchy.

Object composition is a way to combine object(s) into another object, implying ownership. In this paper, the concept of composition has been used to show how the database of hospital is composed of the data from the databases of its departments, as shown in the figure. It simplifies the process of retrieving summary data as well as detailed data from each level of granularity.

In the above Figure 5, we have shown the class diagram of our case study and Figure 6 shows the corresponding snowflake model by using the mapping rules.

Conclusion

This paper presents feasible integration architecture for integrating UML data sources in order to build a multidimensional model for OLAP. Mapping rules have been defined that convert the semantics of UML Schemas into UML Class diagrams in order to extract the multidimensional information from multiple UML data sources. The concept of data granularity models the level of detail present in the data sources. Lower the level of detail, finer be the data granularity. In future, we plan to build a cost model for the integration of UML data sources and evaluate query optimization issues involved in our integration framework.

REFERENCES

1. Abelló A, Samos J., and. Salter F. "A Framework for the Classification and Description of Multidimensional Data Models". 12th Intl. Conference on Database and Expert Systems Applications (DEXA'01), pages 668_677, Munich, Germany,in September 2001.,
2. Alaskar, K. and Shaikh, A." Object Oriented Data Modeling for Data Warehousing (An Extension of UML approach to study Hajj pilgrim's private tour as a Case Study" Bharati Vidyapeeth University, Institute of Management Kolhapur, India IEEE in 2009.
3. Booch,G.,Rumbaugh,J.,and Jacobson,I."The Unified Modelling Language User Guide,"Addison-Wesley Longman, p.482 in 1999.
4. Golfarelli, M., Maio, D., and Rizzi, S."The Dimensional Fact Model: A Conceptual Model for Data Warehouses." Int. Journal of Cooperative Information Systems (IJCIS).7.(2-3): 215-247in 1998.

5. Inmon W., Hackathorn, R. "Using the data warehouse". Wiley-QED Publishing, Somerset,NJ, USA in 1994.
6. Moody, D.L., Kortink, M.A.R. "From Enterprise Models to Dimensional Models : A Methodology for Data Warehouse and Data Mart Design", 2nd International Workshop on Design and Management of Data Warehouses (DMDW 2000), Stockholm,Sweden in June 2000.
7. Rifaiea, M. , Kianmehrb, K. , Alhajjb, R "Data Warehouse Architecture and Design," 978-1-4244-2660-7/08/\$25.00 ©2008 IEEE
8. Sapia, C., Blaschka, M., Höfling, G., and Dinter, "Extending the E/R Model for the Multidimensional Paradigm. " 1st Int. Workshop on Data Warehouse and Data Mining DWDM'98).Singapore: Springer-Verlag LNCS 1998.
9. Sarkar,A.,Choudhury,S.,chaki,N.,bhattacharya S."Conceptual Level Design of Object Oriented Data Warehouse: Graph Semantic Based Model, " National Institute of Technology, Durgapur, West Bengal,India
10. Stefanov, V. and List, B., "A UML Profile for Modeling Data Warehouse Usage"
11. Trujillo, J., Palomar, M., Gómez, J., and Song, I.Y. "Designing Data Warehouses with OO Conceptual Models". IEEE Computer, special issue on Data Warehouses, (34): 66-75 in 2001.
12. Tryfona, N., Busborg, F., and Christiansen, J. "StarER: A Conceptual Model for Data Warehouse Design." ACM 2nd Int. Workshop on Data Warehousing and OLAP DOLAP'99,. Missouri, USA
13. "UML Modeling Language Specification. "March 2009.Version 1.3.
14. Villarroel, R., Fernandez-Medina, E., Trujillo, J., and Piattini, M.. "Towards a UML 2.0/OCL extension for Designing Secure Data Warehouses". in 3rd. Int. Workshop on Security in Information Systems (WOSIS 2005). Miami, USA: INSTICC Press
15. Xuanzi, H. , Kuanfu ,W" Application of Data Warehouse Technology in Data Center Design" Computer Engineering Department,Dongguan City, China. International Conference on Computational Intelligence and Security.

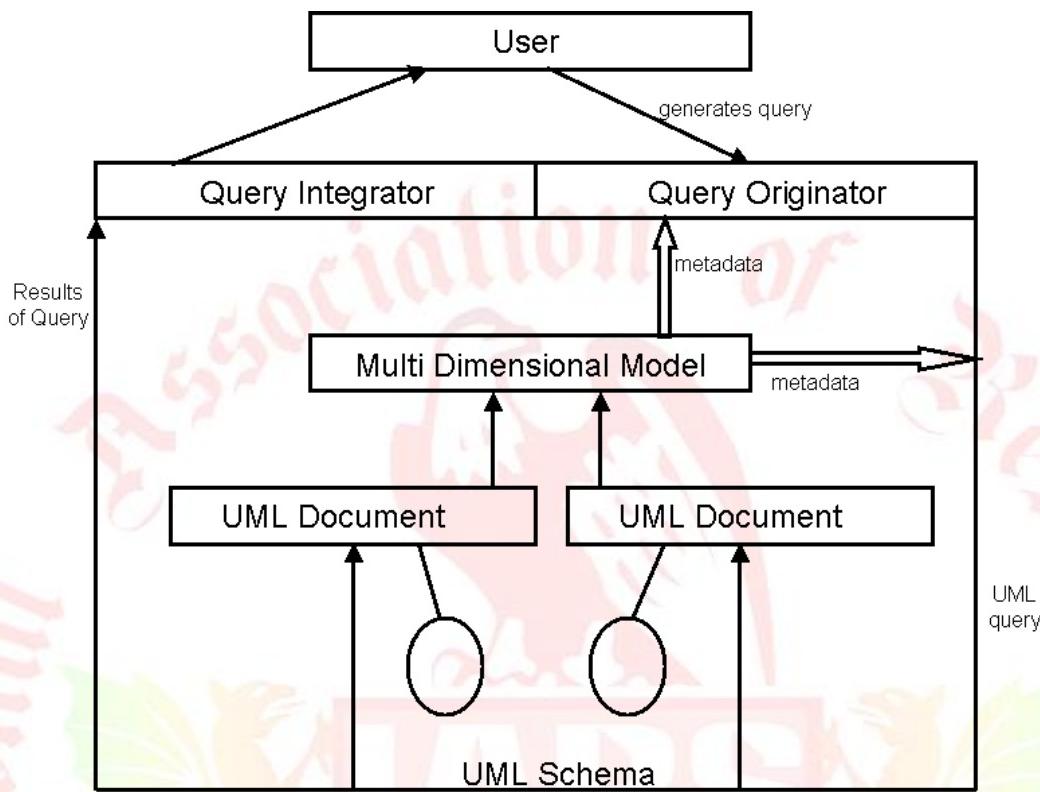


Figure 1: Two -Tier System Architecture

Star/Snowflake Schema of Entire Conceptual Multidimensional Model of D.I.E.T System

Figure 2: Conceptual Model of the system

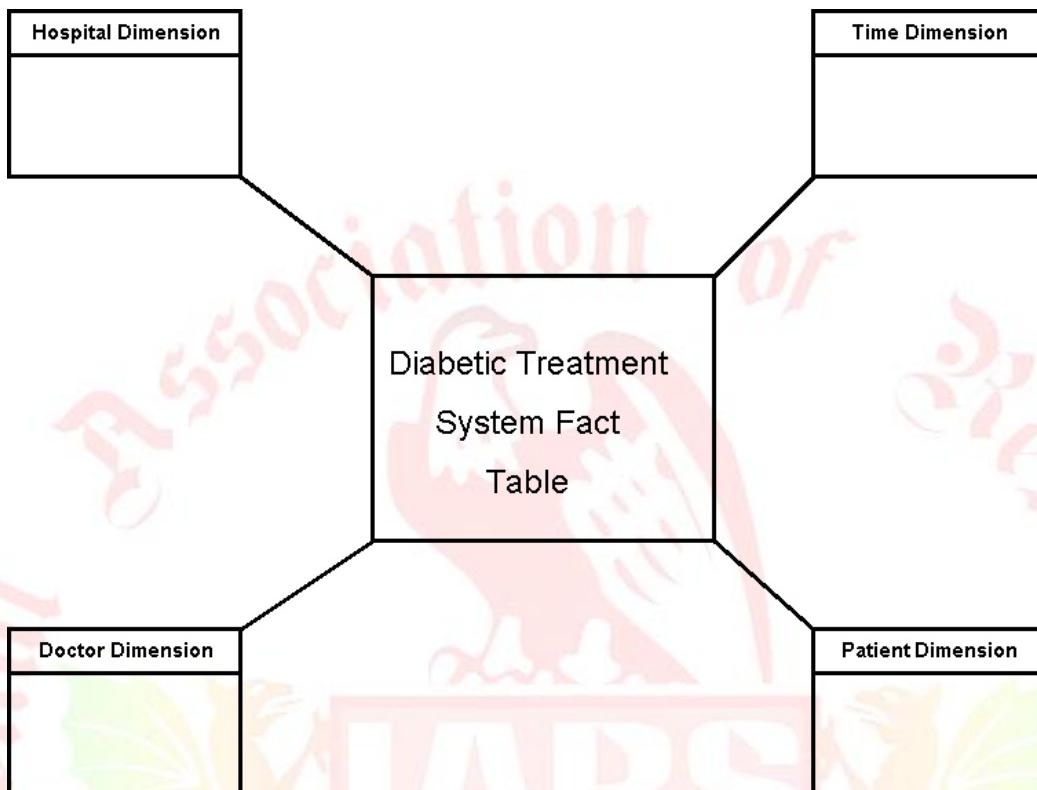


Figure 3: Star Schema of D.I.E.T System

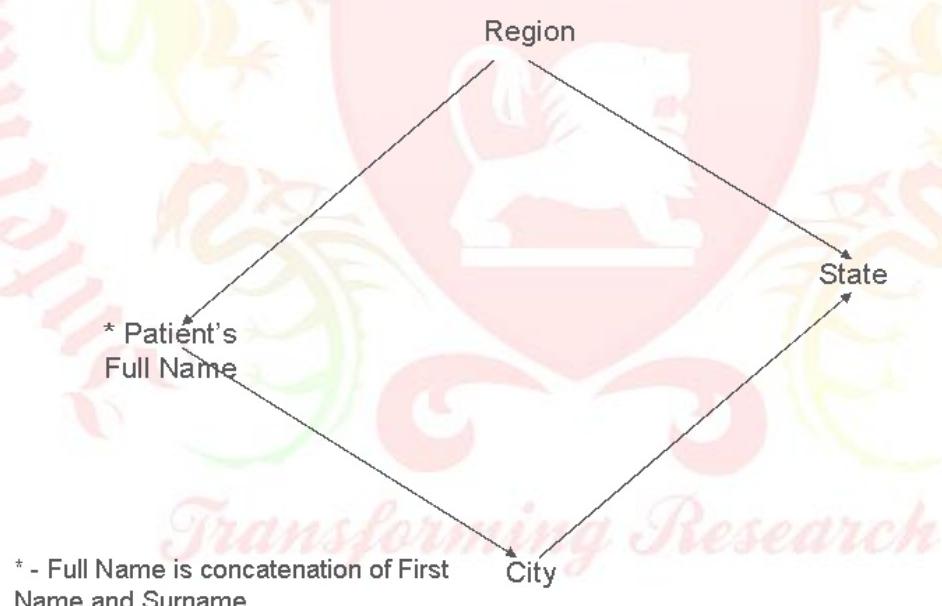


Figure 4: Patient Dimension

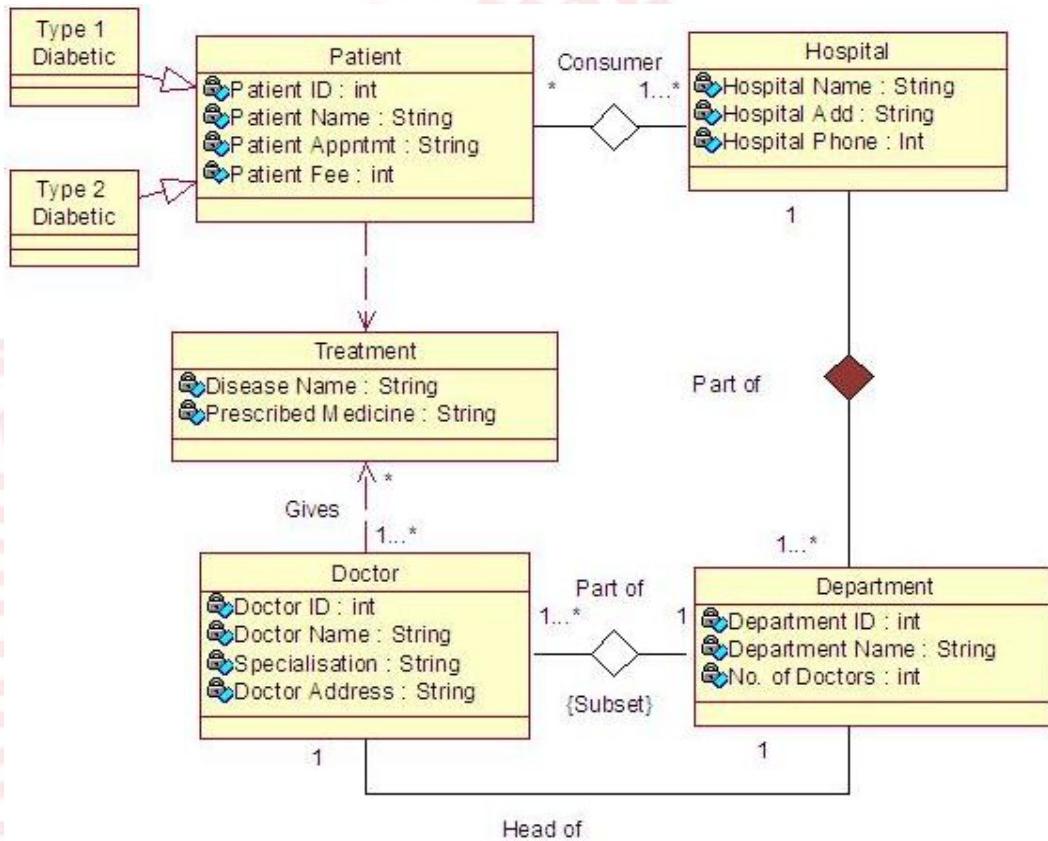


Figure 5: UML class diagram of D.I.E.T

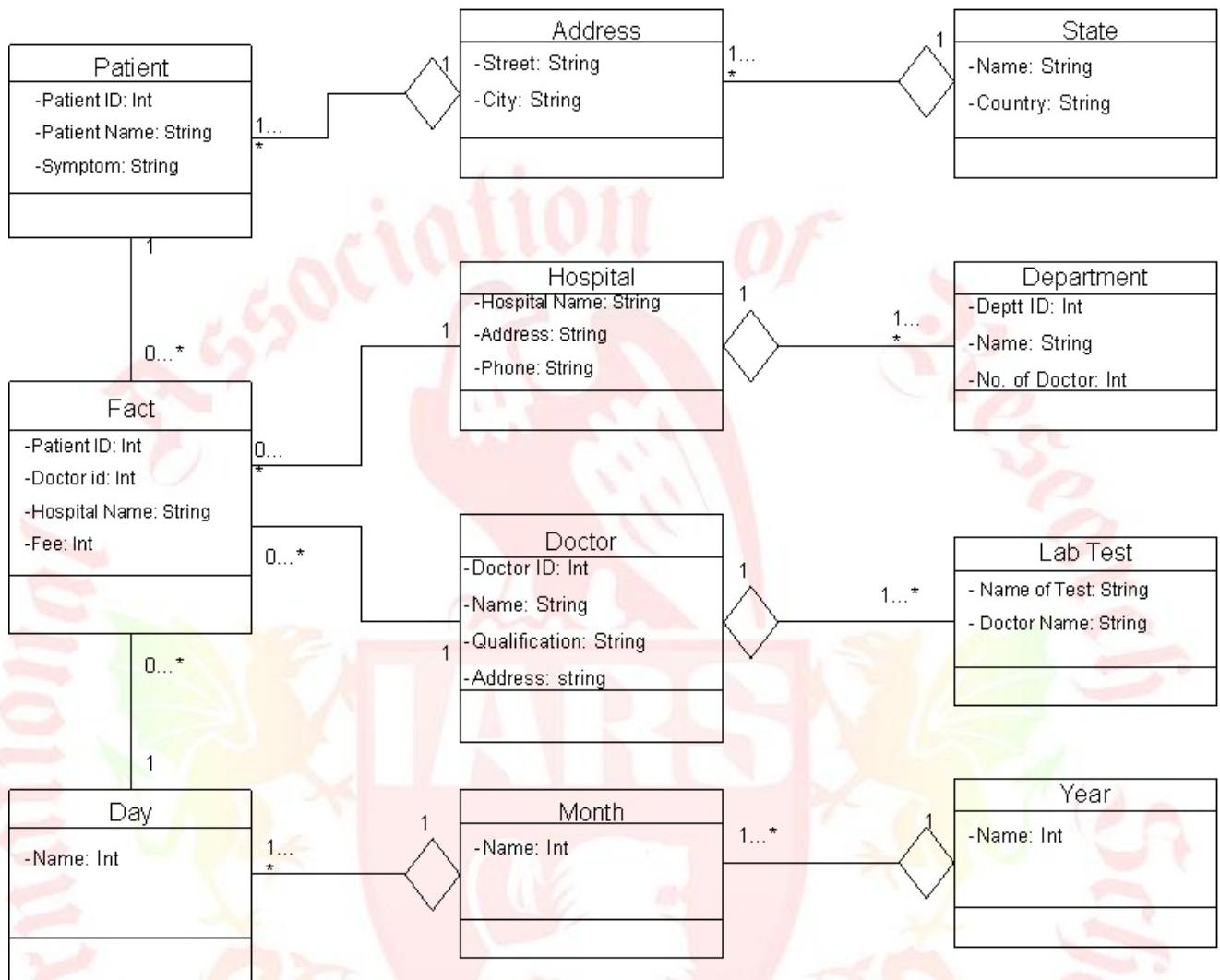


Figure 6: UML Snowflake Model of D.I.E.T

- END -



www.iars.info

Certificate of Recognition

This certificate is awarded to

Shalini Jain

in recognition of his / her contribution

**“Design of a Multidimensional Model Using
Object Oriented Features In UML”**

to Vol. 01, No. 01, 2011 of

IARS International Research Journal
An International Refereed Research Journal

ISSN 1839-6518 (Australian ISSN Agency)



Editor in Chief

